DOCUMENT RESUME

ED 282 890 TM 870 260

AUTHOR Mertens, Sally; Rogers, Glen
TITLE Personnel Issues in Maintaining Longitudinal Designs.
INSTITUTION Alverno Coll., Milwaukee, Wis.
PUB DATE Apr 86
NOTE 18p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, April 16-20, 1986).
PUB TYPE Speeches/Conference Papers (150) -- Viewpoints (120)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Consultants; Data Collection; Labor Turnover; *Longitudinal Studies; *Recruitment; *Research Design; *Researchers; Research Methodology; Research Problems; *Staff Development; Staff Orientation; *Staff Role

ABSTRACT

In longitudinal studies, the same variables must be measured, in the same way, on at least two occasions. Research personnel are very important in keeping the necessary continuity of the procedures and the comparability of the measurements over time. As staff turnover can create serious problems, this paper suggests several strategies for reducing staff attrition in longitudinal studies. When hiring, it should be determined if the candidates are strongly motivated and view the study as related to their future goals. After hiring, staff should be provided with the opportunities which encourage them to make a personal investment in the study. When staff leave, people who understand and accept the conceptual and procedural boundaries of the study should be employed in their place. Prior to the second data collection wave, review of the study's conceptualization by new and continuing staff through the reanalysis of the initial data is recommended. To insure measurement consistency, training current staff through field testing the data collection procedures is suggested. In addition, the importance of establishing procedures for maintaining ongoing written documentation on the instrumentation, coding procedures, and quantitative data files is stressed. Bibliographical information for 10 research reports published by Alverno College are cited on the verso of the title page. (JAZ)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED282890

PERSONNEL ISSUES IN MAINTAINING LONGITUDINAL DESIGNS

Sally Mertens
Glen Rogers

Office of Research and Evaluation
ALVERNO COLLEGE

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

S. Mertens

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Presented at the annual meeting of
The American Educational Research Association
in San Francisco, April, 1986.

TM 870 260

An overview and rationale for our approach to the study of college outcomes, and a summary of the results from the following series of ten research reports, are found in:

Marcia Mentkowski and Austin Doherty. Careering After College: Establishing the Validity of Abilities Learned in College for Later Careering and Professional Performance. Final Report to the National Institute of Education: Overview and Summary. Milwaukee, WI: Alverno Productions, 1983; revised 1984. (ERIC Document Reproduction Service No. ED 239 556.)

Research Reports:

One: Friedman, M., Mentkowski, M., Earley, M., Loacker, G., & Diez, M. Validating Assessment Techniques in an Outcome-Centered Liberal Arts Curriculum: Valuing and Communications Generic Instrument, 1980. (ERIC Document Reproduction Service No. ED 239 557.)

Two: Friedman, M., Mentkowski, M., Deutsch, B., Shovar, M.N., & Allen, Z. Validating Assessment Techniques in an Outcome-Centered Liberal Arts Curriculum: Social Interaction Generic Instrument, 1982. (ERIC Document Reproduction Service No. ED 239 558.)

Three: Assessment Committee/Office of Research and Evaluation. Validating Assessment Techniques in an Outcome-Centered Liberal Arts Curriculum: Insights From the Evaluation and Revision Process, 1980. (ERIC Document Reproduction Service No. ED 239 559.)

Four: Assessment Committee/Office of Research and Evaluation. Validating Assessment Techniques in an Outcome-Centered Liberal Arts Curriculum: Integrated Competence Seminar, 1982. (ERIC Document Reproduction Service No. 239 560.)

Five: Assessment Committee/Office of Research and Evaluation. Validating Assessment Techniques in an Outcome-Centered Liberal Arts Curriculum: Six Performance Characteristics Rating, 1983. (ERIC Document Reproduction Service No. ED 239 561.)

Six: Mentkowski, M., & Strait, M. A Longitudinal Study of Student Change in Cognitive Development, Learning Style, and Generic Abilities in an Outcome-Centered Liberal Arts Curriculum, 1983. (ERIC Document Reproduction Service No. ED 239 562.)

Seven: Much, N., & Mentkowski, M. Student Perspectives on Liberal Learning at Alverno College: Justifying Learning as Relevant to Performance in Personal and Professional Roles, 1982. (ERIC Document Reproduction Service No. ED 239 563.)

Eight: Mentkowski, M., Much, N., & Giencke-Holl, L. Careering After College: Perspectives of Lifelong Learning and Career Development, 1983. (ERIC Document Reproduction Service No. ED 239 564.)

Nine: Mentkowski, M., DeBack, V., Bishop, J., Allen, Z., & Blanton, B. Developing a Professional Competence Model for Nursing Education, 1980. (ERIC Document Reproduction Service No. ED 239 565.)

Ten: Mentkowski, M., O'Brien, K., McEachern, W., & Fowler, D. Developing a Professional Competence Model for Management Education, 1982. (ERIC Document Reproduction Service No. ED 239 566.)

© Copyright 1986. Alverno College Productions, Milwaukee, Wisconsin. All rights reserved under U.S., International and Universal Copyright Conventions. Reproduction in part or whole by any method is prohibited by law.

Personnel Issues in Maintaining Longitudinal Designs

PRECIS

Whereas attrition and substitution of research subjects receive regular attention in the literature, little attention is directed to a parallel problem attending longitudinal studies--namely personnel shifts in the research staff. Personnel shifts are likely to affect most longitudinal studies whether, for example, as a result of funding fluctuations or the natural transientness of staff, many of whom are graduate students. This paper addresses the critical concern of maintaining research design through the delineation of problems and specific examples of strategies for dealing with them. This detail provides the foundation for specific recommendations, especially with regard to training and documentation procedures.

PERSONNEL ISSUES IN MAINTAINING LONGITUDINAL DESIGNS

Although used in many disciplines to study a broad range of questions, the historical home of longitudinal research methodology is in developmental psychology (Baltes & Nesselroade, 1979, p. 8). The longitudinal design, which requires the measurement of the same individuals, on the same variables, on at least two occasions, is typically defined in contrast to the cross-sectional design which entails measurement of groups of individuals of various ages at one point in time. For studies of human development, the longitudinal or repeated measurement methodology is generally viewed as more appropriate than the cross-sectional design in that it clearly delineates "time" as a variable. Thus, a longitudinal design allows for the study of patterns of intraindividual development, of individual as well as group differences in change, and also allows for the examination of links between antecedent and subsequent behavior. There are certainly problems in implementing a longitudinal design. But, if one is interested in the growth patterns of people, Guire and Kowlaski (1979) have gone so far as to maintain that "the only alternative to the longitudinal approach is abandonment of the project" (p. 92).

The most obvious difficulty with the longitudinal design is the time, typically a number of years, required to complete a study. Initiating a longitudinal study requires a tremendous level of commitment in the face of many unknowns regarding the future, long-term capability for conducting the study. In a recent attempt to identify longitudinal studies currently in process, Mednick, Harvey and Finello (1984) identified only 380 studies across all fields. In addition to the time required, there are technical difficulties specific to the analysis of data collected over the years of a longitudinal study. Most thoroughly debated has been the appropriateness of using change scores considering, for instance, the problems of statistical regression, reactivity and practice effects (Cronbach & Furby, 1970; Cook & Campbell, 1979). Attention has also been directed toward developing strategies for dealing with missing data and the attrition of subjects, realities with which every longitudinal study must contend (Rubin, 1974; Goldstein, 1979).

The authors' participation in the fourth data collection wave of a longitudinal study initiated by Alverno College in 1976 has contributed to the development of the views reflected in this paper. The Alverno study of college women provides longitudinal data on Alverno students and alumnae. A description of the findings thus far on a battery of cognitive, motivational, and moral development measures, as well as interviews and questionnaires, are available in the publications listed in the Appendix to this paper. Requests for publications listed in the Appendix should be directed to Alverno Institute, Alverno College, 3401 S. 39th Street, Milwaukee, WI 53215. Requests for this paper may be directed to Sally Mertens, Office of Research and Evaluation, Alverno College, 3401 S. 39th Street, Milwaukee, WI 53215.

At Issue: Comparability of Measurement

In comparison to the attention paid to these technical issues, disproportionately less interest and concern has been directed to what we view to be the heart of the matter, i.e., ensuring the comparability of measures over the span of repeated measurements. The longitudinal design has clear superiority over the cross-sectional approach in that it doesn't have to deal with establishing the comparability of individuals composing the various age groups under study. However, the longitudinal design forfeits this advantage to the extent that comparability of measurement is not maintained over time; the wider the time span under consideration, the more problematic this becomes.

Context for Discussion

As a context for this discussion, let's consider a hypothetical, longitudinal, "life span" study which involves data collection every five years. The initial population included several hundred individuals. Data, collected in the first wave, are available for a multitude of variables. As preparations are being made for initiating the second data collection wave, the following conditions will be assumed: the principal investigators is still committed to the study and is able to provide leadership and direction; the funding level is comparable to that for the first wave and is believed to be adequate for covering the costs of the five full-time staff required for data collection and analysis; the procedures used in the first data collection wave are well documented in technical reports; the computer system used in the first wave is still in operation; raw data are the foundation of the data management system; there is no reason to suspect anything other than random attrition from the subject pool; and, the instruments used in the first wave are still available and are still appropriate. In short, let's assume quite a large-scale study and that conditions are quite optimal for initiating the second wave of data collection.

Focus of Discussion

The longitudinal design requires that the same variables be measured in the same way on at least two occasions in time. In developing a research design, it is generally assumed that the theoretical orientation and the specific questions being addressed will remain constant over the life of the study; it is also assumed that the variables and the procedures for measuring them will remain constant. However, even under optimal conditions, the greater the time span between subsequent data collection waves, the more likely the fundamental premises regarding constancy are likely to be challenged. Therefore, comparability of measurement concerns are much more likely to emerge in longitudinal studies than in other types of studies which are completed within relatively short periods of time.

Much can happen with the passage of time. For example, social change, historical events, theoretical developments, methodological advancements, and developments in computer technology have possible implications for a longitudinal study as do changes in level of funding. Perhaps less obviously, shifts in the personnel conducting the longitudinal study can also have implications.

These staffing shifts are the focus of this paper. We will maintain that comparability of measurement is very much a personnel issue. Each time a staffer leaves the research project, the necessary continuity in the research orientation and procedures is threatened. Each time a new person joins the research staff, this continuity is also likely to be challenged. In short, maintaining the theoretical orientation and the data collection procedures is not as pro forma as might be assumed or hoped. With respect to maintaining the conceptualization of the longitudinal study (i.e., repeated study of the same variables), we will first make suggestions in the areas of staff selection and orientation. With respect to maintaining the same procedures (i.e., variables measured in the same way), we will make recommendations in the areas of staff training and procedures for documentation.

Selection of Professional Staff

Given the mobility of people with the passage of time, the initiation of subsequent data collection waves almost always requires hiring new professional staff. Although staffing shifts attend most large-scale/long-term data collection efforts, these are a particular concern in a longitudinal study since they pose a definite threat to maintaining comparability of measurement. Consider that each new staffer arrives carrying baggage filled, for instance, with a theoretical orientation, special interests, procedural preferences and also probably with some special abilities as well as a few liabilities. In short, new staffers further compound the difficulties inherent in the effort to hold a longitudinal design together over the years. Therefore, every effort should be made to select and retain staff who can make a long-term commitment to the longitudinal study. Every effort should be made to keep staffing shifts to a minimum.

"Matching" the Longitudinal Study and Staff

In hiring, assuming candidates bring a relevant background and the necessary skills (i.e., fit the project), it is important to determine if the candidates view the longitudinal study as related to their future goals (i.e., the project fits the candidates). Although the management literature stresses the importance of determining this latter match (Wanous, 1978; Feldman, 1976), the importance of matching might not be thought much about until the problems, left behind by staff who have gone elsewhere, have to be addressed. The issues in providing continuity and constancy in a longitudinal study are so great, even without turnovers in key personnel, that every effort should be made to select able people who can potentially become committed to the study for a long period of time.

One suggestion to be made is that as part of the screening process, the candidates, having had the opportunity to review the technical reports, be required to write proposals for mini-studies tangential to the longitudinal study. In fulfilling this application requirement, the candidates should arrive at an understanding of both the opportunities and the constraints inherent in becoming affiliated with a longitudinal study. In reviewing the proposals, it should be possible to determine the level of understanding the candidates bring

to what has already been accomplished (i.e., Do the candidates fit the project?). Additionally, this exercise should enable judgements regarding the compatibility of the candidates' orientations to the primary orientation of the ongoing study (i.e., Does the project fit the candidates?). In short, this exercise has the potential of reducing a number of unknown factors for both the principal investigators and the candidates.

Some Insurance Against Staff Attrition

Initially, the opportunity to be part of a major, long-term research effort, to develop important skills, and to work with established researchers (i.e., the kind most likely to receive support for longitudinal work), is likely to operate as a powerful motivator for new staff. However, the allure may be short-term, if the fastidiously, demanding nature of the task is perceived to relate directly to the life work of others, but not to that of oneself. Therefore, for those candidates who make it through the selection process, the principal investigators might consider providing encouragement (and ideally adequate resources) for conducting studies in tandem to the longitudinal study. The threat of staff attrition can probably be reduced if staff have a personal investment in the study (Becker, 1960; Rusbult & Farrell, 1979). Attrition can be reduced if staff view the primary study as providing the sample, context and support for pursuing work they can call their own. Examples of relevant, "independent" studies might include detailed analyses of specific components of the primary data bank and tests of additional hypotheses using existing data. Over the years, in addition to increasing staff commitment, these studies are likely to enhance and strengthen the longitudinal study.

As a side benefit, encouraging the staff to engage in independent studies may dissipate the potentially divisive issue of who "owns" the primary longitudinal data bank. Although, realistically and appropriately, individual staffers should not expect to share greatly in ownership of the primary research, they might be allowed ownership of their own independent contributions. From the perspective of the principal investigators, this would appear to be a low-cost investment with a potentially high pay off with respect to staff commitment.

Although in hiring it makes sense to select individuals with special abilities for specific tasks important to conducting the study, we suggest that once on board every staff person be provided with opportunities to be involved with the various other components of the study. First of all, this approach provides ongoing orientation and sensitization of all staff to the various components of the study, facilitating a better understanding of why certain decisions are made and why it is important to adhere to certain procedures, which may appear arbitrary. This approach should broaden the expertise of each staff member and thus, work as a preventive action against staff boredom and/or burnout. It also serves to strengthen the collective memory of the study. Finally, it provides some insurance for the study in the event of staff attrition, since staff who remain will be able to provide the necessary continuity.

The Consultant Alternative

Although admittedly these suggestions for assimilating new staff make special demands on the principal investigators, the alternatives are not trouble-free either. Uncommitted, professional staff may leave the project as soon as it is personally feasible, leaving the principal investigators to deal with the revolving-door phenomenon which can be very costly in terms of time and energy necessarily diverted to the selection and orientation of new staffers. The option of hiring short-term, professional consultants for specific tasks may avoid staff assimilation problems, but it also has special problems. Working for a set fee and within narrow time frames, external consultants may not have the time to understand the study as thoroughly as might be hoped. Consequently, the specific decisions they may make in performing their tasks can, in some cases, create long-term problems. That a mistake in hiring a consultant has been made may not become apparent until the consultant has been paid and left; (on the other hand, a mistake in hiring staff, may, to at least some extent, be compensated for by training and close supervision). Long-term professional consultants hired on retainer present essentially the same problems as uncommitted, professional staff; they will be inclined to leave their work as soon as a better opportunity appears on the horizon.

This is not to say that there is no role for consultants. Only, that they be used appropriately, i.e., for carefully defined and delimited responsibilities and tasks. For instance, let's say you are exploring the use of Lisrel for the next wave of data analysis and expertise is not available in your current staff. Rather than hiring a consultant to actually perform the computer analysis, money would probably be better spent if a consultant were hired to train the current staff in the new procedures. From the process of training current staff, the consultant is likely to derive a better understanding of project rationales and intentions, and the staff will have learned the new procedures with specific respect to the current task. This approach would seem to be one way of appropriately integrating the special abilities of consultants into a project.

Orientation of Professional Staff

Even assuming the optimal conditions we have established as a context for this discussion, there may be presses for change; and every change has implications for comparability of measurement. The greater the time span, the more one needs to be particularly concerned that the orientations and purposes of research personnel may have changed. Although during the first wave the staff may have shared a consensus on points of interest and emphasis, over time the possibility looms that this consensus has eroded. For instance, some staff may be less interested as a result of new individual priorities. Also, some staff, having had the time to reflect upon the findings of the first wave data analyses, may feel less enthusiasm for the initial research questions. Additionally, staff understanding of the study's history may suffer from even minor gaps in documentation; individual memory may not coalesce with group memory when efforts are made to close these gaps. In short, staff may be scratching their individual and collective heads: "Why did we chose those variables?" or "Why did

we measure them in that way?" If new people have been hired for the project, they will certainly be raising questions about the choice of variables and procedures for measuring them. Regardless of whether the second wave is being initiated with old or new staff, the research conceptualization is likely to be questioned.

Reanalysis of the Initial Data

Prior to initiation of the second data collection wave, we strongly recommend that the current staff thoroughly review the conceptualization which undergirds the longitudinal design. This review should orientate the staff by affirming the relationship between the study's conceptualization and the data sources. Specifically, we recommend that the current staff, with the written procedural documentation as the guide, conduct a reanalysis of the initial data. The primary intent of reanalyzing the initial data is not to confirm the initial data analysis, but rather to establish the current staff's understanding regarding what data are available and their quality.

Assumptions regarding what data are available need to be tested against the realities of the data set. Basic descriptive data should be a primary concern. For instance, are the sampling procedures completely and clearly described? (Or, do the ns vary without clear explanation?) Is complete data available for each subject? (Or, must you be prepared for building upon a base of partial data?) Are individual raw scores available? (Or, are you going to have to live with total scores?) What are the variances for data elicited by the instruments? (Must you be concerned with possible ceiling or floor effects?) Does the prior documentation serve as a reliable and accessible guide to the data set? (Or, must you conduct additional analyses to develop a basic description of the data set?)

It is also advisable to test assumptions regarding the quality of the data. The data set should be treated with the same suspicion Cook and Campbell (1979) have advised with regard to archival data. Are the operational definitions of constructs appropriate and adequate? (Or, is detective work required to determine the label referents?) Have definitions remained constant over time? (Or, do major spurts in "development" cast suspicion on definitional constancy?) Have missing values been identified? (Or, do suspicious regularities in the data suggest that midpoint, average or predicted values have been assigned?) Is there logical consistency where one would expect it in the data set? (Or, must you be prepared to make choices from an array of conflicting data?) Are the internal reliability coefficients for summated measures adequate? (Or, must you be concerned about the conceptual integrity of your instruments?)

In summary, the review of available data should strengthen the base for second wave data collection. The reanalysis of data, conducted with a fresh perspective, can be viewed as a procedural "audit," the importance of which is well articulated in the program evaluation literature (e.g., Guba & Lincoln, 1981). The review may serve to confirm the entire initial design; the measures that are to be repeated have withstood the available tests. In this case, energy can be focused on organizing resources to replicate the initial study. However, it may also illuminate some weaknesses in the first analyses

which should be articulated in later reports. The review may also trigger new ideas for improving or enhancing the study. Whatever the case, most importantly, the current staff, i.e., those responsible for shepherding the study through the second wave, will have had the opportunity to test out their understandings of the available data. The process of approaching the initial data base with specific questions should help to orientate the staff toward the requirements of planning the second data collection wave and the future statistical analyses.

Training Professional Staff

The longitudinal design requires that the same variables be measured in the same way in subsequent data collection waves. Standardization of procedures is not an issue specific to a longitudinal design; it is an issue in all designs. However, these issues are typically exacerbated with the passage of time in a longitudinal study. Obviously, if there are new staff, training in data collection procedures is a paramount priority in a longitudinal design in which maintaining equivalent procedures is such a crucial concern. The training program should also include retraining of continuing staff; it simply cannot be assumed that earlier training from years past is sufficient.

Training for Replication of Procedures

Procedural drift is a major concern in the collection of repeated measures. A general approximation of previous data collection procedures simply is not adequate; every effort needs to be made to replicate the previous procedures as closely as is humanly possible. We believe a formalized, structured training program is best for containing the differential impact of the research staff across data collection waves. The training program should be directed toward staff demonstrating that they can function at a level of precision specified by stated criteria. This may be a particular problem when "old" staff need to be retrained. A natural inclination to be less strict in the second wave training should be acknowledged and resisted. Whatever data collection procedures are being used, staff need to demonstrate that they can create and maintain the same conditions as were present in the first wave.

There are some factors related to training that need to be addressed. A very potent factor which may intrude is a lack of resources to support training, especially lack of time and/or money. Since this is difficult to address once the momentum is pushing toward getting "real" data, the resource issue needs to be addressed in the planning stage as timelines and budgets are being developed. However, even assuming adequate resources, a formal training effort might meet resistance. Training for precise behaviors requires staff acceptance of the authority of procedures and criteria developed at an earlier time by previous staff. In some instances, a special effort may be required to develop the current staff's acceptance.

Field Testing

Prior to initiating the second wave data collection, it is important to establish that the documented procedures are adequate for guiding the collection of comparable data in the second wave. There may be a tendency to believe that since data were elicited once, they will be forthcoming once again. There may be a natural inclination to treat the first wave as a successful field test. These natural tendencies must be resisted. Specifically, it is recommended that all procedures be field-tested with current staff. The field test can be viewed as the culminating orientation and training activity for the current staff.

Field testing procedures in preparation for a subsequent wave in data collection is done, not as a pure test of procedures, but rather to determine if current staff, with prerequisite training, can collect comparable data from the current subject pool using the documented procedures for the first wave. In other words, "improvement" of procedures, in this context, means improving the match between the current procedures and the documented earlier procedures. This improved match, however, may not necessarily correspond with the pure improvement of procedures. For example, it might have been recently established in the literature that it is preferable to administer a certain test in a one-on-one situation. Since group administration was used in the first wave, there is a choice to be made: if you want to keep the measure in the longitudinal strand, then you must readminister it in a group setting. Or, you can switch to the individualized procedures, but then you must conceive of the measure as a component of a different study.

Documentation Procedures

Having had the experience of being very dependent on written documentation in conducting the reanalysis of the first wave data and the field test, the current staff should have become very sensitized to the need for maintaining fastidious records.

Documentation of Start-Up Issues

The staff should have no difficulty understanding the need to document the start-up issues they addressed and how they were resolved. Since the longitudinal or repeated measures design is predicated on the match between the actual procedures across the data collection waves, it is extremely important that the results of the field test be particularly well documented. Was the earlier documentation adequate? If yes, this good fortune should be reported. If no, were interpretations made to close certain gaps in documentation? What alternative strategies were explored? Whatever adjustments were made need to be completely documented to provide guidance to colleagues in subsequent data collection waves.

As a final step before initiating the second wave of data collection, we urge that time be set aside for the staff to document the problems they encountered in the data review and field test. The summary activity should be the development of a list of information

they wish had been available at this point. This list can then be used to develop the formal documentation plan for the second wave. Going through this process should highlight, for the current staff, the need for documentation even at those times in the future when other priorities are certain to make more immediate demands on their time.

Establishing Procedures for Documentation

It seems as though, almost invariably, the need for documentation comes immediately after a major flurry of activity in combination with the immediate need to initiate a new data collection activity. From the context of immediate priorities, it is often difficult to view documentation as being of great importance. Recognizing this, it is crucial to have documentation goals well articulated and to have established very specific processes and procedures for documentation; the more routinized these procedures, the better they are for keeping documentation going even when there are more immediate concerns. For instance, all staffers might be required to keep individual logs of their decisions. Or, weekly meetings might be set aside for staff members to announce, for the written record, all issues that have been raised and how they were resolved. It is important that the quality of the ongoing documentation be regularly monitored. Are all staff participating in documentation? Is the level of detail in the documentation adequate? Is there something in the current phase of the project which suggests a need for additional documentation?

Documentation of Instrumentation

The need for documentation cannot be overstated. In short, if documentation is lacking, there is no basis for making external judgments regarding the extent to which the procedures used in the second wave match those used in the first; there is no foundation on which to base a claim that the repeated measures are comparable. Stated more proactively, document everything you can. The strength of longitudinal findings can be no stronger than the strength of the documentation regarding comparability of measurement.

Because instrumentation slippage is a serious threat to the internal validity of a longitudinal design, the documentation of the actual instrument administration is extremely important. The documentation must clearly specify what instruments or procedures were used, with what people, and under what conditions. For example, the documentation of instrument administration might include, in addition to the written record, a video tape of the actual administration(s). Particular attention needs to be paid to documenting the conditions of the administration that could be having an influence. What were the environmental conditions, e.g., time of day? What were the internal conditions, e.g., participant motivation? If there are any exceptions to the general procedures, these should be clearly identified as such and described.

In addition to documenting conditions from the perspective of the staff, it is also useful to determine the subjects' perceptions of the conditions under which they participated in the study. They might be questioned, for example, with regard to their motivational state, their ability to concentrate, and their memory of earlier data

collection. This provides a check on the current data collection phase; very importantly, if used across the waves, this type of data can be used to substantiate the extent of comparability in the conditions of measurement.

Documentation of Coding Procedures

Whenever data are coded (e.g., categorized, labelled or rated), the potential of "observer drift" exists. There is the potential, irrespective of how rigorously coders were initially trained, that over time individual coders will begin to use criteria other than those according to which they were trained. Central to the prevention of observer drift in coding is a complete set of rules for making inferences. It is important that these rules be based upon sound principles of categorization (see Holsti, 1969); it is simply easier for coders to remember and to use "good" rules.

Observer drift threatens not only test-retest reliability, but also the internal validity of the study. It challenges the very foundation for assuming comparability of measurement (Kratochwill, 1978). Therefore, it is essential that inter-coder agreement be established before any data are coded. Then, it is equally essential that this agreement be checked periodically over the time of data coding; constancy in agreement over time simply cannot be assumed. We believe measures of agreement which adjust for chance, such as Kappa (Fleiss, 1971), provide stronger documentation of inter-coder reliability than calculation of unadjusted percentage agreement.

Documentation of Quantitative Data Files

Thus far, we have stressed the importance of documentation of data collection and coding procedures in providing support for the claim that the measures repeated across time are indeed comparable. A well-organized and documented computer data bank is also important in supporting this claim; conversely, if the data bank is not well documented, the current staff cannot be confident they are accessing data appropriate to procedures for analyzing repeated measures.

Codebooks. The sources for the various data files need to be clearly identified. Codebooks are important for describing the data set, for example, the names of variables, their locations, and their values (see Babbie, 1983). Codebooks provide the link between the data and the computer printouts. Most available statistical packages provide techniques for documenting variables and their values with appropriate labels. Since these labels appear on the printout, effort must be taken to select meaningful labels.

Sample definition. Managing a longitudinal data set, where data are entered at each wave, requires an additional level of documentation. At each wave, attrition and missing data will likely lead to a new definition(s) of "the longitudinal" sample. This redefinition needs to be precisely documented (e.g., Were subjects eliminated?) Are some subjects "longitudinal" on only some of the instruments? Do different research questions require the calling up of different samples? For the sake of clarity and efficiency, we recommend defining a primary longitudinal sample for the bulk of the statistical analyses.

Data transformations. Statistical analysis of repeated measures requires the entry of comparable data. This requires the very precise description of the data; it must be certain that the new data are being meshed appropriately and accurately with the old data. Therefore, scales and indices need to be documented, for instance, in a central file describing all the data transformations and specifying unique names for the created variables. Even if the researcher intends to analyze only transformed variables, it is very important to maintain the raw data set from which the transformations were made. The following hypothetical situation highlights this importance. In the first wave data analysis assume that having computed the sum of two variables, the individual values for each variable were discarded. In the second wave analysis, the two variables are no longer correlated with one another, with the result that the second wave analyst has no justification for summing the variables. There is no way to maintain valid longitudinal measurement, even though the individual variables were themselves valid. To avoid situations such as this, always maintain the raw data sets from which transformations are made.

Listing of data files. When the sample is defined in more than one way or the variables are transformed in more than one way for different analyses, more than one version of the data set may be created. If so, it is important to create a listing of the data files that summarizes the various versions of the data set. For example, a file listing might give the names and a brief description of a file with only total scores for indices, a file with only a subset of the indices, a comprehensive raw data file, and assorted other files.

Dictionary of analyses. Since analyses tend to generate large stacks of computer printouts, it is important to cull and organize. It is helpful to annotate computer printouts with reference to where they appear in reports, and to cross-list reports back to the analyses.

In Summary

The fundamental requirement of the longitudinal design is ensuring comparability of measurement over the time span of repeated measures. With the passage of time there will be many threats to comparability of measurement, many challenges to maintaining the same variables as the focus of study and to measuring them with the same procedures. Documentation is certainly critical in the effort to provide continuity. Succinctly, if documentation is lacking, there is no support for the claim that the repeated measures are comparable. However, one cannot rely totally upon systems for written documentation. As is so often the case in systems of any kind, there is a human factor. In longitudinal studies, there is a need for people who can explain the written documentation and provide information pertinent to closing the gaps in the written record. Because the need for linkage and continuity over time is so central, staff turnover can create especially serious problems in longitudinal studies.

We believe every effort should be made to retain staff (assuming competence) as long as possible. Every instance of staff attrition creates a disruption in the continuity of a longitudinal study which can never be totally compensated for through training of new staff. In the wake of every staff departure, a longitudinal study is likely to suffer from gaps in memory, understanding, and tacit knowledge on routines. Therefore, several strategies for reducing staff attrition have been presented. Central to these strategies is the provision of opportunities which encourage staff to make personal investments in the longitudinal study.

However, even with opportunities and support for individual staff members' personal involvement in the study, people will eventually leave. When this occurs, every effort should be made to hire people who understand and accept the conceptual and procedural boundaries of the study. Although there are many advantages to working on a longitudinal project, there are constraints which should be openly acknowledged. In short, every effort should be made to hire people who match the longitudinal project and to keep them involved in personally meaningful ways for as long as possible.

References

Babbie, E. (1983). The practice of social research. (3rd ed.). Belmont, CA: Wadsworth.

Baltes, P. B., & Nesselroade, J. R. (1979). History and rationale of longitudinal research. In J. Nesselroade & P. Baltes (Eds.), Longitudinal research in the study of behavior and development (pp. 1-39). New York: Academic Press.

Becker, H. S. (1960). Notes on the concept of commitment. American Journal of Sociology, 66, 32-40.

Cook, T. D., & Campbell, D. T. (1979). Quasi-experimentation: Design and analysis issues for field setting. Chicago: Rand McNally College Publishing Co.

Cronbach, L. J., & Furby, L. (1970). How should we measure "change" - or should we? Psychological Bulletin, 74, 68-80.

Feldman, D. C. (1976). A practical program for employee socialization. Organizational Dynamics, 5, 64-80.

Fleiss, J. L. (1971). Measuring nominal scale agreement among raters. Psychological Bulletin, 76, 378-382.

Goldstein, H. (1979). The design and analysis of longitudinal studies. New York: Academic Press.

Guba, E. G., & Lincoln, Y. S. (1981). Effective evaluation: Improving the usefulness of evaluation results through responsive and naturalistic approaches. San Francisco: Jossey-Bass Publishers.

Guire, K. E., & Kowalski, C. J. (1979). Mathematical descriptors and representation of developmental change functions on the intra- and interindividual levels. In J. R. Nesselroade & P. B. Baltes (Eds.), Longitudinal research in the study of behavior and development (pp. 1-39). New York: Academic Press.

Holsti, O. R. (1969). Content analysis for the social sciences and humanities. Reading, MA: Addison-Wesley.

Kratochwill, T. C. (1978). Single subject research. New York: Academic Press.

Mednick, S. A., Harvey, M., & Finello, K. (Eds.). (1984). Handbook of longitudinal research. (Vols. 1 & 2). New York: Praeger Publishers.

Rubin, D. (1974). Characterizing the estimation of parameters in incomplete data patterns. Journal of the American Statistical Association, 69, 467-474.

Rusbult, C. E., & Farrell, D. (1983). A longitudinal test of the investment model: The impact on job satisfaction, job commitment, and turnover of variations in rewards, costs, alternatives, and investments. Journal of Applied Psychology, 68, 429-438.

Wanous, J. P. (1978). Realistic job previews. Personnel Psychology, 31, 249-258.

Appendix

References: Alverno Longitudinal Study

Giencke-Holl, L., Mentkowski, M., Much, N., Mertens, S., & Rogers, G. (1985, April). Evaluating college outcomes through alumnae studies: Measuring post-college learning and abilities. Paper presented at the annual meeting of the American Educational Research Association, Chicago. (ERIC Document Reproduction Service No. ED 261 626)

Mentkowski, M., and Doherty, A. (1984). Abilities that last a lifetime: Outcomes of the Alverno experience. AAHE Bulletin, 36(6), 5, 6, 11-14.

Mentkowski, M., & Doherty, A. (1983; revised 1984). Careering after college: Establishing the validity of abilities learned in college for later careering and professional performance. (Final Report to the National Institute of Education: Overview and Summary.) Milwaukee: Alverno Productions. (ERIC Document Reproduction Service No. ED 239 556)

Mentkowski, M., & Loacker, G. (1985). Assessing and validating the outcomes of college. In P. Ewell (Ed.), Assessing Educational Outcomes. (New directions for institutional research, No. 47). San Francisco: Jossey-Bass.

Mentkowski, M., Moeser, M., & Strait, M. (1983). Using the Perry scheme of intellectual and ethical development as a college outcomes measure: A process and criteria for judging student performance. (Volumes I and II.) Milwaukee: Alverno Productions.

Mentkowski, M., Much, N., & Giencke-Holl, L. (1983). Careering after college: Perspectives on lifelong learning and career development. (Final Report to the National Institute of Education: Research Report No. 8.) Milwaukee: Alverno Productions. (ERIC Document Reproduction Service No. ED 239 564)

Mentkowski, M., & Rogers, G. (1985, October). Longitudinal assessment of critical thinking in college: What measures assess curricular impact? Paper presented at the annual meeting of the Midwest Educational Research Association, Chicago.

Mentkowski, M., & Strait, M. (1983). A longitudinal study of student change in cognitive development, learning styles, and generic abilities in an outcome-centered liberal arts curriculum. (Final Report to the National Institute of Education, Research Report No. 6.) Milwaukee: Alverno Productions. (ERIC Document Reproduction Service No. ED 239 562)